**Cladistic Parsimony, Historical Linguistics, and Cultural Phylogenetics[[1]](#footnote-2)\***

Abstract: Here, I consider the recent application of phylogenetic methods in historical linguistics. After a preliminary survey of one such method, i.e. cladistic parsimony, I respond to two common criticisms of cultural phylogenies: (1) that cultural artifacts cannot be modeled as tree-like because of borrowing across lineages, and (2) that the mechanism of cultural change differs radically from that of biological evolution. I argue that while perhaps (1) remains true for certain cultural artifacts, the nature of language may be such as to side-step this objection. Moreover, I explore the possibility that cladistic parsimony can be justified even if (2) is true by appealing to the inference pattern known among philosophers as ‘Inference to the Best Explanation’ (IBE).

**1. Introduction**

Recently, within historical linguistics, a number of studies attempting to reconstruct the historical relationship between extant languages have been undertaken using methods normally used by biologists to infer evolutionary history. In biology, evolutionary history is often represented on branching, tree-like diagrams known as ‘phylogenetic trees’, or ‘evolutionary trees’, or just ‘trees’ for short. So too in linguistics, language history is often displayed in a similar fashion. Among those phylogenetic methods used by biologists to feature in historical linguistics, one that has proved quite popular is cladistic (or maximum) parsimony, a brief overview of which I’ll provide in section 2. The other prominent method that biologists use to infer the topology of phylogenetic trees is called ‘maximum likelihood.’ This method has also been increasingly applied to language data, about which I’ll have more to say in section 6. Henceforth, I’ll refer collectively to this nascent research program as ‘linguistic phylogenetics’, or more broadly as ‘cultural phylogenetics’ , which includes the use of such methods to study other elements of culture. Despite the prospect of rendering more exact our knowledge of the history of languages, the analysis of language data using phylogenetic methods has not been met with wide acceptance among historical linguists (Nichols and Warnow 2008, p. 760).[[2]](#footnote-3) Two general objections to cultural phylogenetics loom large: (1) the history of cultural artifacts, such as languages, cannot be modeled as tree-like because of borrowing across lineages, and (2) the mechanisms of cultural change, including language change, differ radically from the mechanisms of biological evolution.

In this article, my purpose is twofold. First, I aim to bring these exciting methodological debates to a wider, interdisciplinary audience. Second, I aim to analyze to what extent these general objections undermine linguistic phylogenetics. Before doing that, I begin by explaining how cladistic parsimony works in biology and consider briefly one such parsimony analysis on languages. Here, I focus on cladistic parsimony for a number of reasons: (1) the method is relatively simple and non-technical, and thus serves as an accessible example of phylogenetic inference; (2) the first major objection that I discuss applies equally to all phylogenetic methods, and so a detailed survey of all such methods is not necessary to appreciate this objection; and, most important, (3) the nature of maximum parsimony, but not maximum likelihood, might be such as to allow parsimony to avoid the second major objection. As I discuss in sections 6 and 7, if cladistic parsimony does not depend on its being vindicated by maximum likelihood in order to be justified—as some proponents of parsimony in biology aver—then challenges to the evolutionary models assumed by likelihood methods miss their mark, at least as concerns the use of parsimony. In support of a kind of non-statistical justification of parsimony—a possibility which has gone unappreciated in the methodological reflections of both opponents and proponents of language phylogenies—I propose the novel view that such a defense might naturally find a home in the epistemological framework known among philosophers as ‘Inference to the Best Explanation’.

**2. Some Biological and Cladistic Preliminaries**

In addition to the theory of natural selection, the other great triumph of Darwin’s *On the Origin of Species* is the advancement and defense of the theory of common ancestry. This is the idea that any two organisms, including those that belong to different species, will have, if we look far back enough in time, some ancestor in common from which both are descended. What’s more, it is not only organisms which are morphologically similar, such as coyotes, wolves, and foxes that are related because of their descent from a common ancestor, but rather, Darwin surmised ‘all organic beings which have ever lived on this earth have descended from some one primordial form, into which life was first breathed’ (1859, p. 484). In the last 150 years since Darwin’s bookwas first published, a fruitful research program has succeeded in amassing an abundance of evidence for the truth of the theory of common ancestry. Moreover, recent statistical analyses support the stronger claim that Darwin himself was cautious to assert, namely that there is one universal progenitor of all living things (Theobald 2010).

For contemporary biologists the difficult task that remains is to reconstruct the way in which the tree of life is structured. The theory of common ancestry says that gray wolves, coyotes, and red foxes are all genetically related. But we want to know which two are more closely related, if, for example, wolves and coyotes share a common ancestor that is not at the same time a common ancestor of foxes. Put differently, we want to know if wolves and coyotes form a ‘monophyletic group’, i.e. a group that includes some ancestral organism and all and only its descendants. In this case, there are three distinct possibilities: 1) wolves belong to a group with coyotes that excludes foxes, 2) wolves belong to a group with foxes that excludes coyotes, and 3) foxes belong to a group with coyotes that excludes wolves.

Three distinct possible phylogenetic trees

 One method that biologists have employed to tackle the problem of reconstructing the topology of the tree of life is that of cladistic parsimony. Like other principles of parsimony, such as Ockham's razor, which counsels us not to postulate entities beyond those that are necessary, cladistic parsimony is also concerned with the minimization of some quantity. But instead of minimizing entities, cladistic parsimony counsels us, when constructing trees, to minimize the number of ‘homoplasies’, i.e. the independent reappearance of some given character trait. Of course, this is not to say that homoplasies do not ever occur in nature. A classic example of a homoplastic trait is the ability to fly in bats and birds. Even though both birds and bats can fly, the most recent common ancestor of bats and birds could not fly. On the other hand, the ability to fly in both sparrows and robins is not homoplastic, but is instead ‘homologous’, as the most recent common ancestor of sparrows and robins did have the ability to fly.

 In broad outline, to perform a simple application of cladistic parsimony on the three taxa above, one needs to first choose a set of character traits and then determine which state each of the taxa is in.[[3]](#footnote-4) These traits may be dichotomous, such as the presence or absence of canine teeth, but they need not be. Suppose that we pick 100 character traits and score wolves, coyotes, and foxes accordingly, where a 1 represents the presence of that trait and a 0 represents the absence of that trait. Next, in order to get a parsimony analysis off the ground, it is necessary to determine which traits are ‘plesiomorphic’, i.e. ancestral, and which state is ‘apomorphic’, i.e. derived. This can be done in a number of ways, one of which is by looking at the character states of some taxon which is thought not to belong to the clade whose genealogy is being reconstructed—an ‘outgroup’—and assigning the character states of the outgroup member the plesiomorphic state. Finally, what one needs to do is to determine how many homoplasies each of the three respective tree topologies shown above would require in order to accommodate the observed distributions of the 100 character traits in wolves, coyotes, and foxes.

 On the one hand, certain tree topologies will require homoplasies for certain distributions of traits, and other tree topologies will not, the latter of which are thus favored by those distributions. On the other hand, certain distributions will be uninformative for the reason that these distributions can be accommodated on any of the trees with only one evolutionary change and no homoplasies. In general, for cladistic parsimony, only matchings of traits that are in the derived state, i.e. ‘synapomorphies’, are evidentially relevant, whereas matchings of traits that are in the ancestral state, i.e. ‘symplesiomorphies’, are evidentially irrelevant. In the case at hand, and in general for three taxa, one simply needs to pick the tree with the smallest number of required homoplasies. Of course, few phylogenetic problems are this simple. Because the number of possible topologies increases to 34,459,425 when considering only 10 taxa (Felsenstein 1978b, p. 31), performing a parsimony analysis on even a tiny fraction of the millions of species identified is computationally intractable. Consequently, for more complicated problems, sophisticated computer algorithms have been developed to search for the most parsimonious tree.[[4]](#footnote-5) Biologists are thus forced to rely on the power of computers in order to continue Darwin’s project.

**3. Phylogenetic Methods in Historical Linguistics**

Most of the attempts to apply phylogenetic methods to language data have been concerned with reconstructing the history of major language families, such as the Indo-European (Rexova et al. 2003), Austronesian (Gray and Jordan 2000), Bantu (Holden 2002; Rexova et al. 2006), and Papuan (Dunn et al. 2005) language families.[[5]](#footnote-6) While the method of maximum parsimony has been popular in biology since the late 20th century, in recent years, in light of concerns over how and whether parsimony is justified, new, more complicated methods, such as maximum likelihood and Bayesian approaches, have come to rival parsimony (Steel and Penny 2000, p. 839), especially in studies that use DNA sequences as character traits. In addition to parsimony, these other phylogenetic methods have also been applied to language data (e.g. Gray and Atkinson 2003; Gray, Drummond, and Greenhill 2009; Dunn et al. 2011; Bouckaert et al. 2012) to infer the structure of major language families.

As discussed in the previous section, a parsimony analysis—or any phylogenetic analysis for that matter—requires that one has available a set of character traits on the basis of which one can score the different taxa that are being analyzed. In the phylogenetic analyses that have been done on languages, the characters used have consisted of a variety of linguistic properties. These characters take the form of lexical, morphological, phonological, or syntactical features, or some combination thereof.

An example of a lexical character could be membership in a cognate set associated with the meaning *hand*. To code for this character trait in the Indo-European language family, for instance, one considers the various sets of cognates which mean *hand* in the language family, and then one assigns a language a 1 if it belongs in that cognate set and a 0 if it does not. There may be, and often is, more than one cognate set associated with any given meaning in a language family, and thus more than one character trait associated with that meaning. So, for instance, since *hand* in German is ‘Hand’ and in English is ‘hand’—both of which derive from the Proto-Germanic form ‘\*handuz’ (Skeat 2005, p. 259) –English and German belong to the same cognate set—call it C1—and so receive a 1 for membership in the cognate set C1.[[6]](#footnote-7) In Russian, *hand* is ‘ruká’—which derives from the Proto-Slavic form ‘\*rǫka’ (Barford 2001, p. 18)—and so Russian is not a member of C1, since ‘ruká’ is not a cognate of ‘Hand’ or ‘hand’. Thus, Russian receives a 0 for this character trait. Furthermore, since there is more than one set of cognates for *hand* in the Indo-European language family, one adds another character trait to the data set to account for that fact. In Italian, Spanish, and French *hand* is ‘mano’, ‘mano’, and ‘main’, respectively—all of which derive from the Proto-Italic form ‘\*manus’ (de Vaan 2008, pp. 363-4)—in which case all three languages belong to a different cognate set—call it C2—and so receive a 1 for that character trait. But there is no word for *hand* in Italian, French, and Spanish that is a cognate of ‘hand’ and ‘Hand’. So, unlike English and German, it follows that Italian, French, and Spanish do not belong to C1, and thus these three languages receive a 0 for that trait. Likewise, German and English receive a 0 for membership in the set C2, as there are no cognates of ‘mano’ or ‘main’ in English and German. So too, Russian receives a 0 for membership in C2, as ‘ruká’ is not a cognate of ‘mano’ or ‘main’.[[7]](#footnote-8)

Similarly, a morphological character trait, such as the presence of a conjugated future tensecan also be coded. For such a trait, Italian, Spanish, and French receive a 1 because they all have a conjugated future tense, which derives from their Latinate origin. On the other hand, German, English, and Russian receive a 0 because they lack a future tense, all of which represent the future construction by means of auxiliary verbs. In addition, phonological characters, such as particular sound changes can be coded, and other syntactical/structural features such as the presence of prepositions can also be coded as dichotomous traits. Normally, the coding of linguistic character traits proceeds in this fashion.[[8]](#footnote-9)

 As an example of an application of phylogenetic methods to language data, consider the attempt in Holden (2002) to reconstruct the phylogeny of the Bantu language family, a group of 450 languages spoken across Africa south of the fifth parallel.[[9]](#footnote-10) In this study, a parsimony analysis was run on 73 languages of the Bantu language family, in accordance with available lexical data. In addition, two closely related languages were selected as outgroups based on the likely location of the ancestral language. The data on the basis of which the Bantu tree was constructed includes 92 items of basic vocabulary, such as *man*, *woman*, *tongue*, *fire*, etc., where different cognate sets were treated as different character traits in the manner described above. Search algorithms were used in an attempt to find the shortest tree, and the results consisted of an un-weighted tree with a consistency index of .65, and a weighted tree (weighted on the basis of words thought more likely to change) with a consistency index of .72, which is comparable to biological trees with similar numbers of taxa, suggesting that the language family is largely tree-like.[[10]](#footnote-11)



A portion of the Bantu language tree (Holden 2002, p. 796).

Another crucial feature of the Bantu study is that, according to Holden, it sheds light on controversial questions regarding population and cultural history. In particular, it is consistent with a hypothesis concerning the spread of farming across modern Bantu-speaking Africa. Many researchers who perform phylogenetic analyses on language data attempt to argue for some archaeological or anthropological hypothesis on the basis of their tree constructions. For instance, Gray and Jordan (2000), using maximum parsimony, argue that the reconstructed tree, even with a consistency index of only .25, is evidence that colonization of Polynesia by pre-historic residents of Taiwan must have been relatively rapid. Rexova et al. (2006) suggest an ‘unorthodox scenario of Bantu expansion’ (p. 189) on the basis of a new parsimony analysis performed on more languages and with more characters in addition to those used in the analysis done by Holden (2002).

**4. A Presumptive Argument in Favor of Linguistic Phylogenetics**

Before addressing the controversy surrounding attempts to use phylogenetic methods to infer language trees, it is necessary to consider the motivations for appealing to these methods from biology in the first place. To fully appreciate these motivations, it is necessary to consider first the procedure normally used by historical linguists for establishing language families. This procedure is called the ‘comparative method’. [[11]](#footnote-12) The comparative method is perhaps best illustrated by example, but can be described abstractly as a sequence of steps. In broad outline, in using the comparative method one infers languages families by way of the reconstruction of an ancestral *proto-language*. First, one begins with a set of languages already suspected to be related. Second, one collects a *cognate set*, i.e. a collection of words or morphemes in the languages being investigated which are thought to be related because they descend from an ancestral language.[[12]](#footnote-13) Third, one determines *sound correspondences*, i.e. the sounds found in related words of the cognate set which correspond among the related languages. Fourth, one reconstructs the *proto-phonology*, i.e. the sounds that featured in the proto-language on the basis of the phonetic properties of the daughter languages and ‘conventional wisdom regarding the directions of sound changes’ (Durie and Ross 1996, p. 7). Fifth, one uses the reconstructed proto-phonology to reconstruct *proto-morphemes*. Sixth, one establishes the *shared innovations*, (e.g. phonological, lexical, etc.) of groups of languages relative to the proto-language in order to construct the family tree. Finally, the completion of the project lies in constructing an etymological dictionary for the various languages in the language family, tracing the origin of the words in their respective lexicons.

 The motivations for appealing to phylogenetic methods in historical linguistics are many, and I will mention only some of the most important ones here. On the one hand, a number of researchers have noted apparent parallels between both linguistic and biological evolution, which makes linguistics amenable to biological methods.[[13]](#footnote-14) On its surface, it seems that, just as in the case of organisms, languages descend in a Darwinian fashion from common ancestors. For example, many English-speakers at some point are confronted with the historical fact that unlike French, Spanish, and Italian, which are Romance languages descended from Latin, English is more closely related to modern German. It is thus natural to use the language of ancestry and descent when it comes to describing the history of languages. And when we look closely at languages, say Dutch, English, and German, we notice sufficient overall similarities to suggest relatedness, such as the sound correspondence of the [*t*] in English and Dutch (‘tongue’ and ‘tong’; ‘twelve’ and ‘twaalf’) and the [*ts*] in German (‘Zunge’; ‘zwölf’). This regularity is robust enough to suggest common ancestry (Lass 2003, pp. 52-53). Darwin himself noticed similar linguistic correspondences, and in fact used analogies of cultural evolution in general to elucidate his own proposal of biological evolution (1871, pp. 78-79).

 In addition to the idea that languages already seem ready-made for the methods of biological systematists, another motivation for appealing to phylogenetic methods in linguistics is the inherent limitations of the comparative method. First, researchers commonly cite the claim that there is a certain time in the past for which the comparative method breaks down, usually about 8,000-10,000 years ago. After so extensive a period of time, it becomes impossible to distinguish between genuine inheritance and chance resemblance (Nichols 1992, p. 2). The prospect of capturing a phylogenetic signal from deep into the past, and on the basis of which resolving long-standing historical problems, makes new methods desirable. Second, even though linguists have managed to group languages into broad families, when the inference problem concerns determining which languages in a given family are more closely or distantly related, the task becomes more difficult. As McMahon and McMahon (2005) put it: ‘although we might be able to group languages into a family with a certain measure of security, using the comparative method for instance, subgrouping is still a matter of considerable unclarity, since the method as it stands does not allow for the quantification of degrees of relatedness’ (p. 27). This is a problem because sometimes close linguistic relationships hide themselves, as in the case of the Hindi word ‘čakkā’ and English word ‘wheel’ which, despite all appearances to the contrary, are true cognates derived through a series of regular sound shifts from the Proto-Indo-European reconstruction ‘\*kʷékʷlos’ (Hock and Joseph 2009, p. 440). Finally, despite its prevalence the comparative method has been criticized for lack of rigor and its tendency toward subjectivity (McMahon and McMahon 2003, p. 14). In order to attain a more exact knowledge of the fine structure of language families, these methods need to be supplemented by more rigorous, quantitative methods such as those from biology.

Furthermore, it should be noted that the comparative method already contains similarities to cladistic parsimony, and in fact it has been described as cladistic in nature (Holden 2002, p. 793). For instance, as evidenced in the sixth step of the comparative method, only shared innovations count when delimiting language groups, just as only synapomorphies count, according to parsimony, when delimiting biological clades. Moreover, the ‘conventional wisdom’ featured in step four is sometimes said to include considerations of parsimony. Other things being equal, it is better to postulate a single sound change to account for the distribution of the phonetic properties of daughter languages rather than multiple changes; other things being equal it is better to postulate a syntactic retention from an ancestral language rather than to postulate a change induced by contact with a neighboring language.[[14]](#footnote-15)

 From the foregoing considerations one can derive a kind of presumptive argument in favor of historical linguists exploring phylogenetic methods from biology. Some of the reluctance on the part of linguists to embrace such methods should not be surprising, given that a measure of methodological conservatism is common among practitioners of any field. To some degree, this attitude may be a salutary one to adopt. However, if the current methodology already contains limitations, and if there are new tools which may remedy those problems, and if some of these methods, such as cladistic parsimony, possess deep points of contact with traditional methods, then there seems to be good reason to have confidence in the fruitfulness of these new methods. Thus, unless there is some countervailing reason to think that there are flaws with phylogenetic methods in general, or flaws in their application to language data, then there is a presumption in favor of pursuing linguistic phylogenetics as a research program.[[15]](#footnote-16)

**5. The Problem of Borrowing and Reticulation**

It would be wrong to attribute the skepticism toward linguistic phylogenetics as simple-minded resistance to novelty, however. Much resistance to the idea of extending the phylogenetic methods from biology to the study of human culture derives from the fundamental objection that cultural evolution and biological evolution possess only surface similarities. One serious disanology between organisms and cultures, it is claimed, is that whereas the history of life can be properly represented as a tree, as Darwin does in chapter four of *On the Origin of Species*,the history of culture and cultural artifacts cannot be represented in this way.[[16]](#footnote-17) According to the tree model, each taxon has at most one recent common ancestor, and, while perhaps this is true for biological species, critics argue that this model is inappropriate for describing the history of technological innovations, languages, religious traditions, or cultures in general. This is because there has been far too much ‘reticulation’— i.e., the recombination or amalgamation of lineages— in the history of human culture, owing to the horizontal exchange of cultural information. It is common for cultural traditions to contain salient similarities with their contemporaries, not because of common ancestry, but because of sharing, borrowing, and other interactions. As a result, with respect to cultures or elements of culture, instead of there being only one ancestral source, there may be a half dozen, all of which have contributed in some way to the presently observed characteristics. Thus, the history of human culture looks more like a web or a thicket than a tree. In biological terms, the borrowing and adopting of ideas, beliefs, and even items of vocabulary is analogous to horizontal gene transfer (HGT), whereby genes are transferred from one species to another in any manner other than by sexual reproduction, e.g. by means of a carrier virus. HGT would prove more problematic for biology were the phenomenon ubiquitous, but luckily for cladists, HGT is largely restricted to prokaryotic organisms, despite playing a role in the history of eukaryotes, e.g. in phenomena such as hybridization, endosymbiosis, etc. (Jain et al. 1999, p. 3801).

This objection that cultural evolution is not tree-like was raised in the early 20th century by anthropologist Alfred Kroeber (1948), who argued that with respect to culture:

‘[t]here is a constant branching-out but the branches also grow together again, wholly or partially, all the time. Culture diverges, but it syncretizes and anastomoses too…The tree of culture, on the contrary, is a ramification of such coalescences, assimilations, or acculturations’ (pp. 260-1).

This early charge against modeling the history of culture as a tree retains currency to this day. In an oft-cited passage, biologist Stephen J. Gould essentially repeats Kroeber’s charge:

‘Human cultural evolution proceeds along paths outstandingly different from the ways of genetic change. Biological evolution is constantly diverging; once lineages become separate, they cannot amalgamate…Trees are correct topologies of biological evolution…In human cultural evolution, on the other hand, transmission and anastomosis are rampant’(1987, p. 70).

If the above remarks are correct, then this problem cannot be sidestepped as perhaps it can be in biology, for reticulation is not the exception in the domain of culture, but rather the rule.

 The objection from horizontal diffusion of cultural information is a major challenge to the legitimacy of cultural phylogenetics as a whole. One interesting example where it is alleged that tree-thinking imported from biology gets the history wrong because of reticulation is due to Temkin and Eldredge (2007), who reconstructed using a number of phylogenetic methods, including parsimony, the history of two musical instruments, one of which was the coronet (a brass-wood instrument). Since the history of the development of the coronet is well-documented, this instance provides a vivid case study of the accuracy of phylogenetic methods in the domain of culture. What was found was not favorable to friends of cultural phylogenetics, for the resulting reconstruction of the coronet did not match the known history very well. Owing to a number properties of the history of the coronet, including rampant borrowing of structural features between prior instruments, and even the incorporation of traits from extinct ancestral forms—a type of transmission with no analogue in biology—a tree model where present forms are descended from earlier forms would be historically inaccurate.

This general problem is something of which proponents of linguistic phylogenetics are, of course, aware. For example, Gray and Jordan (2000) write: ‘The issues surrounding the settlement of the Pacific are thus a microcosm of the *general debate* about whether human cultural evolution can be appropriately represented as a tree’ (p. 1053; emphasis mine). Similarly, Holden (2002) writes: ‘Some authors have questioned whether a tree model can describe language evolution...How far relationships among languages are tree-like is part of a *wider debate* on the level of inter-connection between human cultures’ (p. 793; emphasis mine). Unfortunately then, it would seem that the defense of the application of phylogenetic methods, such as cladistic parsimony, in linguistics is in thrall to a foundational debate about whether the evolution of cultures and cultural artifacts sufficiently approximates a tree.

**5.1 Responding to the Objection from Borrowing**

Granted, the example of the incorrect reconstruction of the history of the coronet is a sober reminder of the danger of being too eager to apply new inference methods in places where they don’t belong. However, it would seem that opponents of phylogenetic methods err in the opposite direction. While the history of some aspects of culture may be subject to rampant reticulation, there is no reason to think that this problem is devastating to the project of cultural phylogenetics as a whole, or to the project of linguistic phylogenetics in particular. The idea that reticulation is a major obstacle to the application of phylogenetic methods to matters of culture seems over-stated, given that it is plausibly the case that different domains of culture will differ with respect to the frequency of horizontal transfer and the extent to which such transfer impacts the domain under study.

On the one hand, some elements of culture, such as religious traditions, may be relatively resistant to borrowing, perhaps because of a certain measure of conservatism and early childhood inculcation. Thus, excepting some complications wrought by the occasional phenomenon of religious syncretism, religious traditions may be suitable elements for phylogenetic reconstruction, both independently as a domain unto itself, or as part of a broader cultural phylogeny. On the other hand, other elements of culture, such as technological advances, will not be resistant to borrowing, perhaps because of economic incentives to borrow from past and present lineages. Thus, technological advances will not be suitable elements for phylogenetic reconstruction, neither independently as a domain unto itself, nor as part of a broader cultural phylogeny. However, as far as language phylogenies are concerned, all that needs to be the case is that languages taken in isolation are such as to be resistant to borrowing. One need not assume that cultures in general form relatively stable, coherent units.[[17]](#footnote-18) It might be that few elements of culture have this character, in which case, phylogenetic methods will have a limited application in the study of culture, as many critics have remarked. But, even if that were the case, it would not follow that they have no application whatsoever.

 With that said, the more particular worry that language history is subject to significant borrowing must still be addressed. As it happens, the phenomenon of lexical borrowing is a topic of extensive contemporary research (Haspelmath and Tadmor 2009), and moreover, historical linguists, who hitherto have relied on the comparative method, are well aware of the methodological problems that borrowing poses.[[18]](#footnote-19) As Campbell (1998) emphasizes, ‘[t]he problem of loans and potential loans is very serious ...’ (p. 319).The English language, for instance, serves as a paradigmatic case in point, as 99% of the words in the *Oxford English Dictionary* are loanwords from some other language (Atkinson and Gray 2005, p. 523). Moreover, this problem is potentially compounded by a further difficulty if it turns out that there are absolutely no constraints on contact-induced change, a view which at present is something of an orthodoxy among linguists. It is common to cite a discussion from Thomason and Kaufman (1988) in defense of the complete lack of constraints on linguistic borrowing:

 ‘As far as the strictly linguistic possibilities go, any linguistic feature can be transferred from any language to any other language…This assertion flatly contradicts most older views on the subject and some newer ones as well, but solid evidence has been available and in print for many years’ (p. 14).

If this orthodoxy is correct, then this serves as another significant disanology between languages and organisms. Whereas a biologist can be reasonably sure that despite extended contact between humans and lions, no horizontal gene transfer between these two lineages has occurred, and so a reticulated history need not be a concern, the historical linguist, on the other hand, cannot rest so easily. There is nothing to prevent two extant languages that are radically different from borrowing linguistic features.

 The problem of borrowing in linguistics is a complex one, which, of course cannot be decisively settled here. However, a number of considerations indicate that the problem is not as serious for languages as is sometimes suggested by opponents of cultural phylogenetics. First, according to Heine and Kuteva (2008), the claim that anything may be borrowed from one language to another has been over-stated. There are a number of plausible examples of constraints on contact-induced change, e.g. definite articles are never turned into demonstratives, indefinite articles are never turned into numerals, etc. In addition, to the extent that there are independently defensible laws (or near-laws) of unidirectional change, then such information can be incorporated into phylogenetic reconstructions. Third, genuine hybridizations, i.e. creolizations, are thought to be rare, recent, and atypical. According to Holden and Shennan (2005, p. 25), there are only about ten known, genuine creoles. Thus, a solution to the problem of creoles may be representing such hybridizations as hard polytomies—as a so-called ‘star-phylogeny’—with more than two branches stemming from one node (Harper and Platnick 1978, p. 360). Fourth, contrary to the suggestions of Gould (1987, p. 70), of all aspects of culture, language may be unique in not being prone to five-minute revolutions, such as those effected by the development of tools like the wheel, the stirrup, or the bow and arrow. Rather, owing to the invariant need to communicate and early childhood instruction, language possesses a robust element of conservatism, which in the end may make languages particularly suitable for phylogenetic analysis (Gray et al. 2007, p. 370).

 Furthermore, in most of the phylogenetic analyses hitherto done, at least the ones that code for lexical characters, only a certain collection of words are used, what is often called a ‘Swadesh list’, after the linguist Morris Swadesh, which is a list of 100 or 200 items of basic vocabulary, such as ‘name’, ‘blood’, ‘water’, etc.[[19]](#footnote-20) In many languages throughout the world, it has been shown that these words resist being borrowed, even through linguistic contact because they are used so frequently, unlike, say, words that are used for new technologies, which are greatly subject to borrowing (Pagel et al. 2007; Bowern et al. 2011). While 99% of the words in *Oxford English Dictionary* are the result of borrowing, for the 200-word Swadesh list, that number falls to 6% (Atkinson and Gray 2005, p. 523). Similarly, syntactical features are also thought to be relatively stable and resistant to borrowing. Thus, it is possible to do a phylogenetic analysis with reasonable assurance that the reconstructed tree is not greatly afflicted by the problem of borrowing, provided one uses the right character traits (Dunn et al. 2005). Even though there may be some linguistic diffusion between populations, there nevertheless is good reason to think that languages possess a well-behaved, relatively stable core. While not decisive, the above considerations suffice to show that the problem of reticulation in linguistics is at least not insuperable.

 Contrary to the remarks made by Holden (2002) and Gray and Jordan (2000) at the start of this section then, proponents of phylogenetic methods in linguistics do not need to address wider questions about whether cultures as a whole can be appropriately represented as a tree. The degree of reticulation between cultures is not a debate to be settled globally, but rather must be addressed locally. It is thus unhelpful when critics of cultural phylogenetics put forth general critiques of ‘cladistic theory’, criticizing the application of tree-thinking in such diverse contexts as ethnology, the genealogy of particular technologies, and linguistics all in one fell swoop (e.g. Moore 1994; Terrell 1988). Similarly, it is not clear that we can draw any general conclusions about the nature of cultures or of cultural transmission from the failed recovery of the correct history of the coronet.[[20]](#footnote-21) Instead, in cases where we have reason to believe that borrowing, transfer, and amalgamation of lineages has not occurred, and so the Darwinian assumption of common ancestry is satisfied, then methods like cladistic parsimony may be employed. Importantly for the proponent of phylogenetic methods in linguistics, this assumption is free from wider debates about the coherence and stability of other elements of culture—in which case, they may be defended independently. Finally, new quantitative techniques have recently been applied to cultural data in order to address particular questions of borrowing and branching more rigorously and empirically, which may help to move debates about these issues onto more tractable ground.[[21]](#footnote-22)

**5.2 Phylogenetic Trees and Realism**

One might still object though even if it could be shown that languages are approximately or frequently tree-like. Granted, phylogenetic studies get the correct result that English falls under the Germanic language family, but there is more to English after all than its descent from Proto-Germanic. Borrowings from Latin, Old Norse, and Norman French are part of its history too. But a tree-like model will not represent these relations. So, then, what good is a tree?[[22]](#footnote-23) This worry touches upon the troubling issue of realism with respect to phylogenetic trees in general, a worry which can be directed toward those trees that are constructed by biologists as well. For instance, it’s well known that the very existence of eukaryotic organisms probably owes to the lateral transfer of genes resulting from temporally distant endosymbiotic events, and so a tree of life with three distinct, non-over-lapping domains—Archaea, Bacteria, and Eukarya—will at best under-describe, if not mis-describe the actual history of life.

 The foregoing challenge to language trees can, however, be met by pointing out that, even if no group of languages is perfectly tree-like, representing taxa on hierarchically structured trees provides a useful model for understanding taxonomic diversity in a particular explanatory context. [[23]](#footnote-24) This leaves open whether another model might be more useful for a different problem, or in a different explanatory context. After all, no one model, either tree-like or reticulate, can represent *all* information of interest concerning the history of extant taxa, just as no one map can represent everything that we are interested in mapping. As Kitcher (2001) helpfully puts it, scientists are like map-makers and ‘[m]ap-makers are invariably selective’ (p. 56). Language trees then, like biological trees, show us global, coarse-grained patterns of ancestry and descent; we will have to opt for a different model if we wish to convey horizontal transfer, such as hybridization events or lexical borrowing. If this defense is sound, then it means that language trees, like biological trees, do not tell the whole story; but, it is not as though they don’t convey useful explanatory information, and so surely they shouldn’t be rejected merely because, like almost all scientific models, they contain idealizations and are descriptively incomplete.

**6. Problems of Modeling and Mechanism**

As I have argued, the assumption that the evolution of languages is sufficiently tree-like can, with some qualifications, be defended. But there is another important objection to applying phylogenetic methods to languages that must be addressed. While in the previous section we were concerned with the *pattern* of cultural evolution, in this section we will be concerned with the *process* of cultural evolution. Even if the tree-like assumption is granted, one might object that the mechanism of cultural change is radically unlike that by which organisms evolve. For this reason, cultures and elements of culture are not amenable to investigation through the methods employed by biologists. For instance, one might balk at describing the evolution of Aramaic, Hebrew, and Arabic in terms of natural selection, even if we grant that the three are all daughter languages ‘descended’ from Proto-Semitic. Unfortunately then, it would seem that the success of phylogenetic studies of elements of culture is again in thrall to highly debated, foundational issues about the nature of cultural change. In this case, the issue is whether Darwin’s ideas about how evolution works really constitute, as Dennett (1995) puts it, the ‘universal acid’ (p. 63).

**6.1 Memes, Linguemes, and Natural Selection**

Now, what would be good for friends of cultural phylogenetics would be if the evolution of culture or elements of culture could be shown to fall under one of the abstract theories of natural selection that have been proposed in recent years, such as that of Dawkins (1978) or that of Hull (1980, 1988). In that case, the idea of genuine cultural evolution would not be an illegitimate imposition from biology, but rather a particular application of a general—and, it is hoped, independently plausible—theory of evolution. According to the model of Dawkins, evolution by natural selection occurs when there are such things as ‘replicators’, which refers to any entity that can produce faithful copies of itself, preserving all or most of its features; there is variation among the replicators, which are carried by ‘vehicles’, the generalized notion of an organism; and finally, there is selection for some variants of the replicators but not others. For Dawkins, the vehicles are just along for the ride, as it were. The real locus of selection lies with the replicators. This picture differs from the model of Hull, which accepts the concept of a replicator, but which generalizes the organism as an ‘interactor’—anything that interacts with its environment so as to produce the differential selection of the replicators. If natural selection is just a species of one of these more generalized theories of selection, then in principle there can be genuine cultural evolution, provided that there is some item that fulfills the role of a replicator.

 According to Dawkins that role is fulfilled by what is called a ‘meme’, the concept around which much of the controversy concerning the mechanisms of cultural evolution has turned.[[24]](#footnote-25) A meme is the cultural analogue of a gene: the discrete, bounded unit of inheritance and particular instantiation of the replicator concept in cultural evolution. Putative examples of memes include religious beliefs, ideas, stories, languages, songs, etc. These memes are carried along in the vehicle that is the human mind and are copied on the basis of selective advantage. By the iterated and cumulative process of replication and selection for different memes—it is claimed by Dawkins and his followers—cultures evolve.

 Some proponents of explaining language change by means of the resources of evolutionary theory have taken precisely this approach. For instance, adopting Hull’s generalized theory of selection, Croft (2000, 2002) attempts to explain language change from within the Darwinian framework. Croft uses the term ‘lingueme’ to refer to the ‘unit of linguistic structure, as embodied in particular utterances, that can be inherited in replication’ (2000, p. 239). Linguemes can include physical instantiations of morphemes, phonemes, syntactic rules, words, etc.[[25]](#footnote-26) For Croft, linguemes are the replicators and the speakers of a language are the interactors. Accordingly, the replicators are replicated by means of the speakers’ utterances, and variation is generated by means of interaction with the environment, which consists of other speakers embedded in a particular socio-cultural context.

 Critics of the idea of memes, however, argue that cultural change rarely ever instantiates the pattern schematized by Dawkins or Hull. First, some argue that putative memes are not ever replicated in the way necessary to subsume cultural change under a generalized theory of evolution (e.g. Sperber 2001). In order for an item of culture to be replicated in the relevant sense, the copy must be similar enough to the replicator, the replicator must cause its copy to come into existence, and the information needed for the copy must be obtained from the replicator. Often, however, the last condition is not met. In the case of language acquisition, grammar and vocabulary are not copied by imitation, but rather through more complex processes involving use and inference (Sperber 2001, pp. 169-172). Thus, the mechanism by which cultural artifacts are reproduced seems to be radically un-Darwinian.

A second important objection is to the illicit atomization of culture that results from viewing culture through the lens of replication. A number of critics have argued that cultural artifacts, be they material or abstract, are not properly bounded and discrete in the way required to be replicators (e.g. Bloch 2001, p. 194). Even putative memes that seem like good candidates to be appropriately discrete and bounded turn out to lose their significance when extracted from their cultural and historical context. For example, the belief in God—a favorite example of memeticists—is inextricably tied to practices, such as rituals, or other beliefs, such as the existence of an after-life. So, belief in God cannot be treated as a particulate entity (Kuper 2001, p. 180). Although theories of cultural evolution postulate units of cultural inheritance, Fracchia and Lewontin (1999) put the point boldly: ‘no theory of cultural evolution has provided the elementary properties of these abstract units’ (p. 72). Thus, there may be no properly isolatable unit of culture parallel to the gene, in which case there is nothing for selection to act upon.

The worry for friends of linguistic phylogenetics then is that the criticisms raised toward these approaches to cultural evolution in general will transfer to the evolution of language in particular. First, if critics of the notion of discrete cultural units are right, then it does not seem that we will be able to regard languages or elements of languages coded as character traits in linguistic phylogenetic studies as discrete entities. This assumption, however, seems to be lurking in the background of these studies, about which researchers are often not quite explicit.[[26]](#footnote-27) Second, if critics of the meme concept are right that cultural artifacts are not normally inherited by means of genuine copying of information, then one might think that language evolution in particular will not be able to be assimilated under one of the general theories of evolution by natural selection. This is the view that Andersen (2006) takes, drawing the dismal conclusion that disanalogies in the mechanism of biological evolution and language change are such ‘that there is no chance of explaining language change by the mechanisms of evolutionary theory’ (p. 59). Echoing the objections of Sperber (2001), Andersen argues, among other things, that unlike genetic mutations, language innovation ‘does not involve copying, but abductive inference’, and that whenever an innovation occurs, ‘it is recognizably rational’ (2006, p. 77).

**6.2 A Response to the Objections to Meme-like Cultural Evolution?**

One response sometimes given to these objections to ‘Darwinizing’ culture is to say that a full resolution to the debate surrounding the legitimacy of units of cultural inheritance and the precise mechanism by which they replicate is not necessary (Mesoudi et al. 2004; Gray, Greenhill, and Ross 2007). [[27]](#footnote-28) Because natural selection occurs when there is variation among traits, inheritance from one generation to the next, and differential reproductive success owing to that variation, proponents of cultural phylogenetics who think that cultural change proceeds by a process of selection have argued that one need not specify yet in detail the precise way in which inheritance works in order to be confident that cultures evolve by selection. In a sense, because of its abstract character, the theory of natural selection is neutral with respect to the unit upon which selection acts, as well as the mechanism by which traits are inherited. Indeed, it is this fact that natural selection is substrate-neutral which makes formal, generalized theories of evolution possible. Certainly, we should want to make sure that we properly conceptualize whatever units of selection there are and that we accurately describe the processes of inheritance. But to reject the idea of cultural evolution construed along Darwinian lines because one of the most popular approaches is obscure in matters of inheritance would be akin to mistakenly rejecting Darwin’s own theory at the time he propounded it, since he did not specify in detail the actual mechanism of genetic inheritance.[[28]](#footnote-29)

On top of this, it is important to remember—as Sober (2009, 2011) convincingly argues— that the theory of common ancestry is not only *logically* independent from, but also *evidentially* independent from the theory of natural selection. Darwin himself argued that adaptive similarities tend to conceal rather than reveal common ancestry, as organisms that have no close genealogical relationship may develop similar features owing to common selective pressures in their respective ancestral environments (1859, p. 427). Instead, only instances in which different organisms share traits that are neutral or deleterious will evidence of common ancestry reveal itself.[[29]](#footnote-30) So, far from requiring natural selection, in some sense the presence of natural selection makes questions of common ancestry more difficult to establish. What’s more, it is not only the case that common ancestry does not depend evidentially on natural selection, but also, the theory of common ancestry is evidentially *prior* to the theory of natural selection. In fact, the theory of common ancestry was used by Darwin in the service of arguments for particular hypotheses concerning natural selection, such as the evolution of the eye in vertebrates. More generally, it turns out that facts about natural selection only become testable when viewed through the framework afforded by the theory of common ancestry (Sober 2009, p. 10059). Questions of common ancestry are that which concern phylogenists, and this is distinct from questions concerning the mechanism of evolution, whether that mechanism be natural selection or something else. Thus, given the way in which natural selection and common ancestry are evidentially related, it should be possible to establish that the common ancestry hypothesis is true of a group of cultural artifacts, e.g. a group of languages, independently of establishing the presence of natural selection acting upon that group.

However, an appeal to the evidential independence of the theory of common ancestry from that of natural selection to meet the foregoing objections will only go so far. While it might be possible to establish that the hypothesis of common ancestry is true of a group of languages independently of demonstrating that those languages are subject to the linguistic analogue of natural selection, in order to infer anything substantive about the historical pattern of ancestry and descent, it seems that substantive assumptions about the evolutionary process will be needed. In the context of cladistic parsimony, this requirement surfaces when we consider what would justify an appeal to parsimony as a tree-selection method. After all, one might wonder, why should we be parsimonious with our trees?[[30]](#footnote-31)

**6.3 Likelihood, Parsimony, and Evolutionary Models**

One promising way to justify the use of parsimony would be to show that the most parsimonious tree is also the one that confers the highest probability on the observed characteristics of extant taxa, that is, to show that maximum parsimony agrees with the other prominent method by which biologists infer phylogenetic trees, namely ‘maximum likelihood’. The likelihood of a hypothesis is represented abstractly by Pr(D|H), where H is the hypothesis and D is the data. Choosing the tree with the highest likelihood is straightforwardly justified by the so-called ‘Law of Likelihood’ (Hacking 1965), according to which the hypothesis with the higher likelihood is more favored by the evidence. However, likelihood methods require the assumption of a mathematical model of the evolutionary process, which consists of an abstract, probabilistic description of evolutionary change, and an estimation of the values of the adjustable parameters of that model, e.g. the probability per unit time of a change from state 0 to state 1 on branch X, the probability per unit time of a change from state 1 to state 0 on branch Y, etc. For a language phylogeny, this may include the probability that a new cognate will appear on a certain branch, e.g. a cognate belonging to the cognate set containing ‘main’appears on the Germanic branch. A simple model would make it such that the probability of any change on any part of the tree is the same (e.g. Jukes and Cantor 1969), but more complex models with more adjustable parameter are also possible. Although their justification is more obvious, likelihood methods are much harder to implement because they demand much more by way of specification. A tree topology by itself is not sufficient to confer probabilities on observed character states; only hypotheses of the form ‘Tree topology T & evolutionary model M fitted with values for parameters of M’ have well-defined likelihoods (Sober 2004, pp. 644-9).

Some theorists have sought to show under what conditions the most parsimonious tree will be at the same time the tree with the highest likelihood. Felsenstein (1973) proved that for any two trees T1 and T2 and set of observed character states O, where T1is more parsimonious than T2, Pr(O|T1) > Pr(O|T2), provided that one assumes a model according to which the probability of a character state transformation is low. In addition, Tuffley and Steel (1997) proved the result that for any two trees T1 and T2, where T1is more parsimonious than T2, Pr(O|T1) > Pr(O|T2), provided that one assumes a more complex ‘no common mechanism’ model, according to which the probability of a character state change need not be low but is independent of any other character state change. Thus, this model assumes that each trait evolves by genetic drift. Both of these results suffice to show, under distinct models of the evolutionary process, that maximum likelihood and maximum parsimony give the same result.

However, the problem is that neither of these two models which would suffice to justify the use of parsimony is a realistic general description of the evolutionary process. As Felsenstein himself notes, the assumption that change is improbable is untenable (1973, p. 244). A similar criticism could also be leveled against the model proposed by Tuffley and Steel, concerning the realism of some its assumptions. [[31]](#footnote-32) In any case, it seems that the initial objection to importing phylogenetic methods into historical linguistics resurfaces. What reason is there to suppose that the mathematical models of the evolutionary process used to justify parsimony analyses, or more generally, those which have been used directly in likelihood (and Bayesian) phylogenetic studies, such as those cited in section 2, reflect the real processes that govern language change? For instance, the Tuffley-and-Steel model—which is used in the language phylogeny of Warnow et al. (2006)—assumes neutral evolution, but this is at odds with the fact that some sort of selection seems to operate on languages (Bowern and Evans 2014, p. 10). [[32]](#footnote-33)

Indeed, linguists have pointed out this problem with the stochastic models assumed by Bayesian and likelihood studies. Bowern and Evans (2014) remark that, when it comes to evolutionary studies of language, one ‘weakness in the field at present is the way in which models have been adopted wholesale, without careful attention to consistency in their features’ and ‘without much explicit discussion of the entailments of each model for language data’ (p. 6). As Nichols and Warnow (2008) note, ‘Phylogeny estimation methods based on unrealistic models of language evolution are unlikely to produce accurate estimations of evolutionary history’ (p. 769). Thus, a good deal of the support for language phylogenies depends on the legitimacy of the underlying models which are in question. This problem might be solved if linguists could agree on a general theory of language change—assuming such a theory is possible— something which hitherto has remained elusive (Bowern and Evans 2014, p. 2). One solution to the objections to the process models imported from biology suggested by proponents of linguistic phylogenetics is simply to construct more realistic models of language change (Barbançon et al. 2013). Until that time though, we must be cautious about the prospects of linguistic phylogenetics.

**7. Can There Be a Non-Statistical Justification of Cladistic Parsimony?**

However, another route to pursue, at least when it comes to the application of cladistic parsimony, is to challenge the assumption that parsimony must be justified in terms of its agreement with maximum likelihood and is thereby indirectly wedded to a mathematical model of the evolutionary process. One of the purported virtues of parsimony, so its most ardent proponents say, is that it does not need to make assumptions about the underlying processes by which the transmission of traits is governed (Wiley 1975; Eldredge and Cracraft 1980; Farris 1983; Kluge and Grant 2006). Rather, according to cladists who defend maximum parsimony, the only thing that needs to be assumed when using parsimony in biology is that Darwin’s theory of common ancestry is true. These theorists make a clear distinction between the *pattern* of the tree of life and the *process* by which that pattern came about. When merely inferring the pattern of ancestry, with parsimony at least, detailed information about the nature of the evolutionary process—which for the more complex mathematical models requires the estimation of a large number of ‘nuisance’ parameters—is not necessary, or so it is claimed.

**7.1 Parsimony and Explanatory Power**

Recently, Kluge and Grant (2006) defend this view that there is a justification of the use of maximum parsimony ‘the evolutionary assumptions of which amount to no more than ‘descent, with modification’’ (p. 282). Kluge and Grant offer a non-statistical justification of parsimony, according to which the use of parsimony is licensed by appeal to explanatory power, where the explanatory power of a hypothesis is maximized by minimizing the number of postulated transformation events required to causally explain the distribution of character states of extant taxa. This defense of maximum parsimony is similar to that of Farris (1983), who has adopted the strong position that ‘the modeling approach to phylogenetic inference was wrong from the start’ (p. 17), arguing instead that parsimony is justified by an appeal to explanatory power, where explanatory power is maximized by minimizing ad hoc hypotheses of homoplasy. [[33]](#footnote-34) One might rightly wonder what exactly is meant here by ‘explanatory power.’ While it is natural to think that insofar as Pr(E|H1) > Pr(E|H2), H1 ‘explains’ E better than H2, and thus has more explanatory power, of course Grant, Kluge, and Farris can’t cash out explanatory power directly in terms of likelihoods. Instead, Farris (1983, 2008) and Kluge and Grant (2006) adopt Popper’s measure of explanatory power—call it EPP—, according to which EPP(H, E) = [Pr(E|H)−Pr(E)]/[Pr(E|H)+Pr(E)].[[34]](#footnote-35)

For Kluge and Grant in particular, an explanation E1 is a better explanation than E2 if E1 postulates fewer causal entities than E2 to account for the same phenomenon. This principle of quantitative ontological parsimony, which has been defended recently by philosophers such as Nolan (1997), Barnes (2000), and Baker (2003), says that an explanation E1 has more explanatory power, than explanation E2 , if E1 postulates fewer causal entities than E2 to account for the data.[[35]](#footnote-36) In the matter of inferring phylogenies, ‘explanation is achieved by causally relating the character-states observed in multiple species through their shared transformation events’ (Kluge and Grant 2006, p. 284). So, the tree topology that postulates fewer transformation events is the one that we ought rationally to prefer. This justification, it is claimed, is independent of its agreement with the verdict of likelihood methods. Of course, one need not adopt so strong a view as that of Farris, who seems to think that the modeling approach should be entirely abandoned. Likelihood methods will be legitimate if the models are sufficiently faithful to reality. But, if successful, the preceding justification would vindicate maximum parsimony without resorting to additional, sometimes unrealistic, modeling assumptions.

**7.2 Away from Popper and toward Inference to the Best Explanation**

Likelihood and Bayesian approaches to inferring phylogenies naturally have a home in the Bayesian philosophy of scientific inference (e.g. Howson and Urbach 2006). Even though likelihood methods do not attempt to calculate the posterior probability of different tree topologies, it’s an elementary theorem of the probability calculus that Pr(H1|E) > Pr(H2|E) if and only if Pr(E|H1)∙Pr(H1) > Pr(E|H2)∙Pr(H2), and so all else being equal, higher likelihoods will lead to higher posterior probabilities. On the other hand, champions of cladistic parsimony often attempt to locate their non-statistical defenses of parsimony in Popper’s (1959) falsificationist framework. Given the limitations of Popper’s philosophy of science though—such criticisms I will not rehearse here—an alternative philosophical framework in which to locate non-statistical approaches to defending maximum parsimony is desirable. This is especially pressing if we are to give sense to the appeals made by these cladists to notions of explanatory power.

One such framework I advocate is that of ‘Inference to the Best Explanation’ (IBE), first introduced under this name by Harman (1965), and further defended by Thagard (1978), Psillos (2002), and Lipton (2004), etc. The central idea of IBE is that explanatory goodness is a guide to inference about what is true. IBE is typically formalized as a four-step inference schema, such as the following:

1. F is some fact or collection of facts that requires an explanation.
2. Hypothesis H1, if true, would explain F sufficiently well.
3. No competing explanations H2, H3,...Hn would explain F better than H1.
4. Therefore, one is justified in believing that H1 is true over its competitors.

IBE has the virtue of seeming to capture the nature of a considerable portion of commonsense, scientific, and philosophical reasoning. In a passage oft-quoted by proponents of IBE (e.g. Thagard 1978; Okasha 2000; and Haig 2014) from the end of *On the Origin of Species*, Darwin regards his theory as justified on the basis of explanatory considerations, and furthermore, endorses something like IBE:

‘It can hardly be supposed that a false theory would explain, in so satisfactory a manner as does the theory of natural selection, the several large classes of facts above specified. It has recently been objected that this is an unsafe method of arguing; but it is a method used in judging of the common events of life, and has often been used by the greatest natural philosophers’ (1859, p. 476).

Much more can be said about each of the above four steps than I can say here, but for the purpose at hand, the step of my formulation of IBE that is most relevant is the third. [[36]](#footnote-37)

Most proponents of IBE propose that we ought to rank the success of an explanation by reference to various so-called ‘theoretical virtues’. Proponents of IBE differ with respect to what belongs on the list of theoretical virtues, although there is much overlap, and at times the differences are merely terminological. Quine and Ullian (1970, ch. 6) list the theoretical virtues as including *conservatism, modesty*, *simplicity*, *generality*, and *refutability*.[[37]](#footnote-38) Thagard (1978) defends three standards of evaluation: *consilience*, *simplicity*, and *analogy*. Psillos (2002) lists and discusses the virtues of *consilience*, *completeness*, *importance*, *parsimony*, *unification*, and *precision*. Lipton (2004) cites many of the same virtues listed above, including *mechanism*, *precision*, *scope*, *simplicity*, *fertility* or *fruitfulness*, and *fit with background belief*. Common to all of these lists is simplicity, which no doubt includes the sort of principle of quantitative parsimony to which Kluge and Grant (2006), as well as Farris (1983) appeal. Such defenses of maximum parsimony by appeal to the explanatory power afforded by respecting ontological parsimony thus find a natural home in the IBE framework.

There is a further reason for such non-statistical defenses of cladistic parsimony to separate from Popper’s philosophy of science, namely that Popper’s measure of explanatory power runs up against a number of problems. We can more easily see these problems by first noting that Popper’s measure of explanatory power is ordinally equivalent to the syntactically simpler measure put forth by Good (1960) and McGrew (2003)—call it EPG—according to which EPG = Pr(E|H)/Pr(E).[[38]](#footnote-39) Obviously, EPG trivially entails that some hypothesis H, which simply restates the observations is the most powerful explanation of E, as then Pr(E|H)=1. Similarly, any hypothesis H, which entails E will have a likelihood of unity, but mere entailment, as philosophers have long pointed out, does not always suffice to explain (e.g. Bromberger 1966). Moreover, vacuous explanations, such as an explanation of the drowsy effect of opium consumption in terms of its dormitive virtue, will have a high likelihood and thus count as good explanations according to EPG, but, of course, such hypotheses are hardly explanatory. Finally, there is the problem that tacking onto E any irrelevant conjunct E\* which is probabilistically independent of H given E does not change the value of EPG (Schupbach and Sprenger 2011, pp. 114-5). According to EPG then, to whatever extent Einstein’s general theory of relativity explains the anomalous precession of the perihelion of Mercury, it also explains the conjunction of that fact and the fact that Barack Obama is the 44th president of the United States, which is of course counterintuitive. Since EPG is ordinally equivalent to Popper’s measure EPP, any problems that afflict EPG, also afflict EPP. These problems are sufficient reason for cladists who cling to Popper’s philosophy to abandon’s his measure of explanatory power.

Perhaps some other probabilistic measure of explanatory power that lacks these problems would be well-suited to explicate the notion of explanatory power that defenders of maximum parsimony have in mind. But given the extent to which IBE suitably accommodates non-statistical defenses of maximum parsimony, an analysis of explanatory power in terms of the degree to which the explanation exemplifies the various theoretical virtues is that which cladists should adopt as their working notion of explanatory power. Unlike the probabilistic measure of explanatory power proposed by Popper, proponents of IBE measure the power of an explanation in terms of a plurality of distinct virtues. Explanatory power is constituted by how much understanding an explanation provides, and how much understanding it provides is a function of how well the explanation exemplifies the various theoretical virtues. Normally, proponents of IBE do not specify any formal measure of explanatory power, [[39]](#footnote-40) but presumably such a function—call it EPI— would be, if not linear, at least strictly increasing. Whether each virtue counts the same in EPI, or whether, say, simplicity should be given less weight than predictive accuracy is an issue that we need not address here. Even without the formal details of this measure of explanatory power, given the picture offered by IBE, we know enough to deduce that, for any two competing hypotheses H1 and H2—say these are tree topologies—and evidence E, if H1 and H2 are alike in all respects, except that H1 is more ontologically parsimonious than H2, then EPI (H1, E) > EPI (H2, E).

**7.3 Likelihood, Parsimony, and Historical Linguistics**

While some likelihoodists, e.g. Felsenstein (1987), might view the foregoing considerations as ‘merely’ philosophical, it hardly needs mentioning that likelihoodism is no less a philosophy of science than is IBE. This is especially worth mentioning given that the use of maximum parsimony to infer language phylogenies has recently been criticized on dubious grounds as an argument in favor of using maximum likelihood. Dunn (2014), who is a proponent of language phylogenies, appeals to the problem of statistical inconsistency, and in particular, the phenomenon of long-branch attraction, as a deficiency of maximum parsimony, recommending the use of likelihood methods instead, which he remarks are ‘not subject to this problem’ (p. 197). However, we know that even maximum likelihood can be inconsistent, and what’s more, simulations suggest that, just like maximum parsimony, even maximum likelihood can be subject to long-branch effects (Kuck et al. 2012). Thus, likelihood methods afford no advantage over maximum parsimony, at least with respect to the problem of statistical inconsistency. [[40]](#footnote-41)

Intricate discussion of the relationship between and comparative advantages of parsimony and likelihood is, of course, not possible here. Even so, the idea that the justification of maximum parsimony might not depend on the Law of Likelihood, and would thus not require assuming a probabilistic model of the evolutionary process, is a view which has gone unappreciated in the debate over the adequacy of phylogenetic approaches to the study of language and culture, and is certainly one worthy of further exploration. Whether the minimalist view, according to which the *only* assumption that parsimony makes is that evolution occurred, is correct remains unclear, as much is yet to be uncovered about the assumptions underlying maximum parsimony (Sober 2004, p. 651).[[41]](#footnote-42) Concerning the legitimacy of maximum parsimony in historical linguistics, a lot may in the end depend on certain other philosophical and methodological commitments, whether one thinks, for instance, that an inference method must always be capable of being vindicated by some further statistical justification.

Nevertheless, those linguists who are skeptical of the transference of existing evolutionary models to language data might wish to align themselves with those cladists who try to articulate non-statistical justifications of parsimony. Given the methodological points of contact raised in section 4 with respect to the comparative method and cladistic parsimony, this seems a desirable alliance. If a justification of parsimony could be given that does not require the assumption of a stochastic model, which many linguists have found objectionable, then *some* of the grounds of skepticism among linguists regarding the applicability of phylogenetic methods would be undermined. As I’ve argue, IBE provides a useful framework in which to think about this possible justification of parsimony and better accommodates the sort of non-statistical defense that might be mounted in favor of cladistic parsimony than the Popperian framework. If such a defense is successful, then not all phylogenetic methods would be subject to the second major objection posed by linguists and anthropologists concerned about the inadequacy of the modeling assumptions used in phylogenetic studies.

**8. Conclusion**

Here, I have surveyed the recent explosion of historical studies of language that employ methods that were originally developed in biology for the purposes of inferring evolutionary lineages. As I have argued, there are a number of reasons for pursuing linguistic phylogenetics as a research program, both internal and external to historical linguistics. The two major objections to linguistics phylogenetics, I have argued, are not insuperable. First, I have suggested that the issue of horizontal transmission across cultures is something that must be addressed locally, rather than globally. Even if reticulation is on the whole rampant in the history of human culture, there is much reason to think that the problem, at least as concerns linguistics, is less serious than previously supposed. In the end, a model which represents both vertical and horizontal transmission may be most accurate, but it does not follow from this that a tree-like model is thereby rendered useless. Second, while it is true that the objection stemming from the inadequacy of the process models assumed by maximum likelihood and Bayesian analyses constitutes a legitimate worry, as I have maintained, both proponents and opponents of cultural phylogenetics have failed to consider the possibility that cladistic parsimony might be justified independently of its agreement with maximum likelihood. Following proponents of cladistic parsimony in biology, I have suggested that parsimony might be defended non-statistically by appealing to considerations of explanatory power, and in particular, I have proposed that such a defense might be better accommodated and explicated by Inference to the Best Explanation. Although to fully meet the second objection to linguistic phylogenetics more realistic models of language change would need to be developed, nonetheless, if successful, the defense of phylogenetic analyses of language that I have sketched here undermines some of the grounds for skepticism toward linguistic phylogenetics.

Department of Philosophy

University of Wisconsin-Madison

References

Aikhenvald, A.Y. and Dixon, R.M.W. (eds.) 2001: *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. Oxford: Oxford University Press.

Akaike, H. 1973: Information Theory and an Extension of the Maximum Likelihood Principle. In B.N. Petrov and F.Csaki (eds.), *2nd International Symposium on Information Theory.* Budapest: Akademiai Kiado, 267‑81.

Andersen, H. 2006: Synchrony, Diachrony, and Evolution. *Current Issues in Linguistic Theory,* 279, 59-90.

Atkinson, Q.D., and Gray, R.D. 2005: Curious parallels and curious connections—phylogenetic thinking in biology and historical linguistics. *Systematic Biology*, 54(4), 513-526.

Aunger, R., (ed.) 2001: *Darwinizing Culture: The Status of Memetics As a Science*. Oxford: Oxford University Press.

Baker, A. 2003: Quantitative parsimony and explanatory power. *The British Journal for the Philosophy of Science*, 54, 245–259.

Barbançon, F., Warnow, T., Evans, S., Ringe, D. and Nakhleh, L. 2013: An experimental study comparing linguistic phylogenetic reconstruction methods. *Diachronica* 30(2), 143-170.

Barford, P.M. 2001: *The Early Slavs: Culture and Society in Early Medieval Eastern Europe*. Ithaca, NY: Cornell University Press.

Barnes, E.C. 2000: Ockham’s Razor and the anti-superfluity principle. *Erkenntnis*, 53, 353–374.

Baum, D. and Smith, S. 2012: *Tree Thinking: An Introduction to Phylogenetic Biology*. Greenwood Village, CO: Roberts and Company Publishers.

Bloch, M. 2001: A Well-disposed Social Anthropologist’s Problems with Memes. In R. Aunger (ed.), *Darwinizing Culture: The Status of Memetics As a Science*, 189-204.

Bouckaert, R. et al. 2012: Mapping the origins and expansion of the Indo-European language family. *Science*, 337: 957-60.

Bowern, C., Epps, P., Gray, R., Hill, J. Hunley, K., McConvell, P., and Zentz, J. 2011: Does Lateral Transmission Obscure Inheritance in Hunter-Gatherer Languages?, *PloS One* 6.9.e25195.

Bowern, C. and Evans, B. 2014: Editors' introduction: foundations of the new historical linguistics. In Bowern, C. and Evans, B. (eds.), 2014: *The Routledge Handbook of Historical Linguistics*, 1-42.

Bowern, C. and Evans, B. (eds.), 2014: *The Routledge Handbook of Historical Linguistics*. Abingdon: Routledge.

Boyd, R., Borgerhoﬀ-Mulder, M., Durham, W.H., Richerson, P.J. 1997: Are cultural phylogenies possible? In P. Richerson, P.J. Mitchell, S.D. Maasen, and S.Mahwah (eds.), *Human by Nature, Between Biology and the Social Sciences*. NJ: Lawrence Erlbaum, 355-386.

Bromberger, S. 1966: ‘Why Questions’. In R. Colodny (ed.), *Mind and Cosmos: Essays in*

*Contemporary Science and* Philosophy. Pittsburgh: University of Pittsburgh Press, 86-111.

Brössel, P. 2013: Correlation and Truth. In V. Karakostas and D. Dieks (eds.), *Recent Progress in Philosophy of Science: Perspectives and Foundational Problems*. Dordrecht: Springer.

Cabrera, F. 2015: Can there be a Bayesian explanationism? On the prospects of a productive partnership, *Synthese*, 1-28. doi: 10.1007/s11229-015-0990-z

Camin, J.H. and Sokal, R.R. 1965: A Method for Deducing Branching Sequences in Phylogeny.
*Evolution* 19(3), 311-326.

Campbell, L. 1998: *Historical Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

Collard, M. and Tehrani, J. 2005: Cladistic analysis of Turkmen textiles sheds light on cultural

evolution. In R. Mace, R., C.J. Holden and S.J. Shennan (eds.), *The Evolution of Cultural Diversity: A Phylogenetic Approach*, 109-131.

1. Croft, W. 2000: *Explaining Language Change : An Evolutionary Approach.* London: Longman.
2. Croft, W. 2002: The Darwinization of Linguistics. *Selection*, 3, 75-91.

Darwin, C. 1859: *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Darwin, C. 1868: *The Variation of Animals and Plants under Domestication (Vol. 2)*. London: John Murray.

Darwin, C. 1871: *The Descent of Man, And Selection in Relation to Sex*. London: John Murray.

Dawkins, R. 1978: Replicator selection and the extended phenotype. *Zeitschrift fur Tierpsychologie*, 47, 61–76.

Dennett, D. 1995: *Darwin’s Dangerous Idea*. USA: Simon & Schuster.

Dixon, R.M.W. 1997: *The Rise and Fall of Languages*. Cambridge, UK: Cambridge University Press.

Donoghue, M.J. 1985: A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist*, 88(3), 172-181.

1. Doolittle, W.F. and Bapteste, E. 2007: Pattern pluralism and the Tree of Life hypothesis, *Proceedings of the National Academy of Sciences*, 104(7), 2043-2049.
2. Dunn, M. 2014: Language phylogenies. In C. Bowern, & B. Evans (eds.), *The Routledge Handbook of Historical Linguistics*. Abingdon: Routledge, 190-211.

Dunn, M., Greenhill, S. Levinson, S., Gray, R.D. 2011: Evolved structure of language shows lineage-specific trends in word-order universals. Nature 473, 79–82.

Dunn, M., Terrill, A., Reesink, G., Foley, R.A., and Levinson, S.C. 2005: Structural Phylogenetics and the Reconstruction of Ancient Language History. *Science*, 309(5743), 2072-2075.

Durie, M. and Ross, M., 1996: Introduction. In M. Durie and M. Ross (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, 3-38.

Durie, M. and Ross, M., (eds.) 1996: *The Comparative Method Reviewed: Regularity and Irregularity in Language Change.* Oxford: Oxford University Press.

Eldredge, N. and Cracraft, J. 1980: *Phylogenetic Patterns and the Evolutionary Process*. New York, NY: Columbia University Press.

Farris, J.S. 1983: The Logical Basis of Phylogenetic Analysis. In N. Platnick and V. Funk (eds.), *Advances in Cladistics - Proceedings of the 2nd Annual Meeting of the Willi Hennig Society*. New York, NY: Columbia University Press, 7-36.

Farris, J.S. 1999: Likelihood and inconsistency. *Cladistics* 15, 199-204

Farris, J.S. 2008: Parsimony and Explanatory Power. *Cladistics*, 24, 1-23.

Felsenstein, J. 1973. Maximum likelihood and minimum-step methods for estimating evolutionary trees from data on discrete characters. *Systematic Zoology*, 22, 240-249.

Felsenstein, J. 1978a: Cases in which parsimony and compatibility methods can be positively misleading. *Systematic Zoology*, 27, 401-410.

Felsenstein, J. 1978b: The Number of Evolutionary Trees. *Systematic Zoology*, 27(1), 27-33.

Felsenstein, J. 1987: Comment (on Statistical analysis of hominoid molecular evolution). *Statistical Science*, 2, 208–209.

Forster, M.R. and Sober, E. 1994:  How to Tell when Simpler, More Unified, or Less *Ad Hoc*Theories will Provide More Accurate Predictions. *British Journal for the Philosophy of Science* 45**,** 1-35.

Fracchia, J. and Lewontin, R.C. 1999: Does Culture Evolve? *History and Theory*, 38(4), 52-78.

Glymour, C. 2015: Probability and the Explanatory Virtues. *Brit. J. Phil. Sci*., 66(3), 591-604.

Goloboff, P.A. 2003: Parsimony, Likelihood, and Simplicity. *Cladistics*, 19, 91-103.

Good, I.J. 1960: Weight of Evidence, Corroboration, Explanatory Power, Information and the Utility of

Experiments. *Journal of the Royal Statistical Society*, *Series B (Methodological)*, 22 (2),319-331.

Gould, S.J. 1987: *An Urchin in the Storm: Essays about Books and Ideas.* New York: WW Norton & Company.

Gray, R.D. and Jordan, F.M. 2000: Language trees support the express-train sequence of Austronesian expansion. *Nature*, 405, 1052–55.

Gray, R.D. and Atkinson, Q.D. 2003: Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, 426, 435-439.

Gray, R.D., Drummond, A.J., and Greenhill, S.J. 2009: Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science,* 323(5913), 479-83.

Gray, R.D., Greenhill, S.J., Ross, R.M. 2007: The Pleasures and Perils of Darwinizing Culture (with Phylogenies), *Biological Theory*, 2(4), 360-375.

Gray, R.D., Bryant, D., and Greenhill, S.J. 2010: On the shape and fabric of human history. *Philosophical Transactions of the Royal Society* *B*, 365, 3923-.3933

Greenberg, J. 1977: *A New Invitation to Linguistics*. Garden City, NY: Anchor Books.

Greenhill, S.J. and Gray, R.D. 2005: Testing Population Dispersal Hypotheses: Pacific Settlement, Phylogenetic Trees and Austronesian Languages. In R. Mace, C.J. Holden, S. Shennan *The Evolution of Cultural Diversity: A Phylogenetic Approach*, 31-52.

Hacking, I. 1965: *The Logic of Statistical Inference*. Cambridge: Cambridge University Press.

Haig, B. 2014: *Investigating the Psychological World: Scientific Method in the Behavioral Sciences*. Cambridge, MA: MIT Press.

Harman, G. 1965: The Inference to the Best Explanation.  *Philosophical Review*, 74, 88-95.

Harper, C.W., and Platnick, N.I. 1978: Phylogenetic and cladistic hypotheses: a debate, *Systematic Biology*, 27(3), 354-362.

Haspelmath, M. and Tadmor, U. (eds.) 2009: *Loanwords in the World's Languages: A Comparative Handbook*. Berlin: De Gruyter Mouton.

Hauser, M.D, N. Chomsky, and Fitch, W.T. 2002: The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598), 1569-79.

Heine, B., and Kuteva, T. 2008: Constraints on Contact-induced Linguistic Change, *Journal of Language contact*, 2(1), 57-90.

Hennig, W. 1966: *Phylogenetic Systematics.* Urbana, IL: University of Illinois Press.

Henrich, J., Boyd, R., and Richerson, P.J. 2008: Five Misunderstandings about Cultural Evolution *Human Nature*, 19(2), 119-137.

Hock, H.H., and Joseph, B.D. 2009: *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Berlin: Walter de Gruyter.

Holden, C.J. 2002: Bantu Language Trees Reflect the Spread of Farming across Sub-Saharan Africa: A Maximum-Parsimony Analysis, *Proceedings of the Royal Society of London. Series B: Biological Sciences*,269(1493),793-799.

Holden, C.J. and Shennan, S. 2005: Introduction to Part I: How Tree-like is Cultural Evolution?. In R. Mace, C.J. Holden, S. Shennan (eds.), *The Evolution of Cultural Diversity: A Phylogenetic Approach*. Walnut Creek, US: Left Coast Press, 1-10.

Holder, M.T., Lewis, P.O. and Swofford, D.L. 2010: The Akaike Information Criterion will not choose the No Common Mechanism model. *Systematic Biology* 59(4), 477-485.

Howson, C. and Urbach, P. 2006: *Scientific Reasoning: The Bayesian Approach*. Peru, IL: Open Court.

Hull, D.L. 1980: Individuality and Selection, *Annual Review of Ecology and Systematics,*11, 311–32.

Hull, D.L. 1988: *Science as a Process: An Evolutionary Account of the Social and Conceptual*

*Development of Science*. Chicago: University of Chicago Press.

Jain, R., Rivera, M.C., and Lake, J.A. 1999: Horizontal Gene Transfer among Genomes: The

Complexity Hypothesis, *Proceedings of the National Academy of Sciences,* 96(7), 3801-3806.

Janda, R. and Joseph B. 2003: On Language, Change, and Language Change – Or, Of History, Linguistics, and Historical Linguistics. In B. Joseph and R. Janda (eds.), *The Handbook of Historical Linguistics*, 3-180.

Joseph, B. and Janda, R. 2003: *The Handbook of Historical Linguistics*. Oxford: Wiley-Blackwell.

Jukes, T. and Cantor, C. 1969: Evolution of protein molecules. In H. Munro (ed.), *Mammalian protein metabolism.* New York, NY: Academic Press, 21-132.

Kitcher, P. 2003: *Science, truth, and democracy*. Oxford: Oxford University Press.

Kluge, A.G. and Grant, T. 2006: From conviction to anti-superfluity: old and new justifications of parsimony in phylogenetic inference, *Cladistics,* 22, 276–288.

1. Kroeber, A.L. 1948: *Anthropology: Race, Language, Culture, Psychology, Pre-history.* New York, NY: Harcourt, Brace and Co.
2. Kuck P., Mayer C., Wagele J.W., and Misof B. 2012: Long branch effects distort maximum likelihood phylogenies in simulations despite selection of the correct model. PLoS One 7: e36593.
3. Kuhn, T. 1977: Objectivity, Value Judgment, and Theory Choice. In *The Essential Tension*. Chicago: University of Chicago Press, 320-39.

Kuper, A. 2000: If Memes Are the Answer, What is the Question?. In R. Aunger (ed.), *Darwinizing Culture: The Status of Memetics As a Science*. Oxford: Oxford University Press, 175-188.

Lass, R. 1990: Early mainland residues in southern Hiberno-English, *Irish University Review,* 20(1), 137-148.

Lass, R. 1997: *Historical Linguistics and Language Change*. Cambridge: Cambridge University Press.

Lass, R. 2003: Genetic Metaphor in Historical Linguistics, *Alternation,* 10(1), 47-62.

Lipton, P. 2004: *Inference to the Best Explanation*, *2nd ed.* New York: Routledge.

Mace, R., and Holden, C.J. 2005: A Phylogenetic Approach to Cultural Evolution, *Trends in Ecology & Evolution,* 20(3), 116-121.

Mace, R., Holden, C. J., Shennan, S., (eds.) 2005: *The Evolution of Cultural Diversity: A Phylogenetic Approach.* Walnut Creek, US: Left Coast Press.

Mayr, E. 1942: *Systematics and the origin of species from the viewpoint of a zoologist*. New York: Columbia University Press.

McGrew, T. 2003: Confirmation, heuristics, and explanatory reasoning. *British Journal for the Philosophy of Science,* 54(4), 553-567.

McMahon, A. and McMahon, R. 2003: Finding Families: Quantitative Methods in Language Classification, *Transactions of the Philological Society,* 101(1), 7-55.

McMahon, A. and McMahon, R. 2005: *Language Classification by Numbers*. Oxford: Oxford University Press.

Mesoudi, A., Whiten, A., and Laland, K.N. 2004: Perspective: Is Human Cultural Evolution Darwinian? Evidence Reviewed from the Perspective of The Origin of Species, *Evolution,* 58(1): 1-11.

Mishler, B.D., and Brandon, R.N. 1987: Individuality, pluralism, and the phylogenetic species concept, *Biology and Philosophy* 2(4), 397-414.

Moore, J.H. 1994: Putting Anthropology Back Together Again: The Ethnogenetic Critique of Cladistic Theory, *American Anthropologist*, 96(4), 925-948.

Nichols, J. 1992: *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.

Nichols, J. and Warnow, T. 2008: Tutorial on Computational Linguistic Phylogeny, *Language and*

*Linguistics Compass* 2(5), 760-820.

Nolan, D. 1997: Quantitative Parsimony. *British Journal for the Philosophy of Science* 48, 329-43.

Okasha, S. 2000: Van Fraassen’s Critique of Inference to the Best Explanation. *Studies in the History and Philosophy of Science*, 31, 691-710.

Pagel, M., Atkinson, Q.D., and Meade, A. 2007: Frequency of Word-use Predicts Rates of Lexical Evolution throughout Indo-European History, *Nature*, 449(7163), 717-720.

Pawley, A.and Syder, F. 1983: Natural selection in syntax: Notes on adaptive variation and change in vernacular and literary grammar, *Journal of Pragmatics,* 7, 551–579.

Popper, K. 1959: *The Logic of Scientific Discovery*. London: Hutchinson.

Psillos, S. 2002: Simply the Best: A Case for Abduction. In A.C. Kakas and F. Sadri (eds.), *Computational Logic: Logic Programming and Beyond*. Berlin: Springer-Verlag, 605-26.

Quine, W.V.O. and Ullian, J. 1978: *The Web of Belief*. New York: Random House.

Rankin, R. 2003: The Comparative Method. In B. Joseph and R. Janda (eds.), *The Handbook of Historical Linguistics.* Oxford: Wiley-Blackwell, 183-212.

Rexova, K., Frynta D., Zrzavy J. 2003: Cladistic analysis of languages: Indo-European classification based on lexicostatistical data, *Cladistics,* 19, 120–27.

Rexova, K., Batin, Y. and Frynta, D. 2006: Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data, *Naturwissenschaften,* 93, 189-194.

Ringe, D., Warnow, T. and Taylor, A. 2002: Indo-European Computational Linguistics, *Transactions of the Philological Society,* 100(1), 59-129.

Roche, W. and Sober, E. 2013: Explanatoriness is evidentially irrelevant, or inference to the best explanation meets Bayesian confirmation theory. *Analysis*, 73, 659-668.

Schupbach, J. and Sprenger, J. 2011: The Logic of Explanatory Power. *Philosophy of Science* 78,105-27

Skeat, W. 2005: *An Etymological Dictionary of the English Language.* Mineola, NY: Dover Publications, Inc.

Sober, E. 1988: *Reconstructing the Past: Parsimony, Evolution, and Inference*. Cambridge, MA: MIT Press.

Sober, E. 2004: The Contest between Parsimony and Likelihood. *Systematic Biology* 53, 644-653.

Sober, E. 2009: Did Darwin Write the Origin Backwards? *Proceedings of the National Academy of Sciences,* 106(Supplement 1), 10048-10055.

Sober, E. 2011: *Did Darwin Write the Origin Backwards?* Amherst, NY: Prometheus Books.

1. Sober, E. 2015: *Ockham's Razors: A User's Manual*. Cambridge: Cambridge University Press.

Sperber, D. 2001: An Objection to the Memetic Approach to Culture. In R. Aunger (ed.), *Darwinizing*

*Culture: The Status of Memetics As a Science*,163-173.

Steel, M. and Penny, D. 2000: Parsimony, Likelihood, and the Role of Models in Molecular Phylogenetics *Mol. Biol. Evol.,* 17(6), 839–850.

Sterelny, K. 2006: Memes Revisited. *The British Journal for the Philosophy of Science,* 57(1), 145-165.

Swadesh, M. 1952: Lexico-statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos, *Proceedings of the American Philosophical Society,* 96(4), 452-463.

Swadesh, M. 1955: Towards greater accuracy in lexicostatistic dating, *International Journal of American Linguistics,* 21, 121-137.

Temkin, I., and Eldredge, N. 2007: Phylogenetics and Material Cultural Evolution. *Current Anthropology,* 48(1), 146-154.

Terrell, J.E. 1988: History as a family tree, history as an entangled bank: constructing images and interpretations of prehistory in the South Pacific. *Antiquity*, 62, 642-657.

Thagard, P. 1978: The Best Explanation: Criteria for Theory Choice. *The Journal of Philosophy*, 75(2), 76-92.

Theobald, D.L. 2010: A Formal Test of the Theory of Universal Common Ancestry, *Nature,* 465: 219-222.

Thomason, S.G., and Kaufman,T. 1991: *Language Contact, Creolization, and Genetic Linguistics.* California: University of California Press.

Towner, M.C., Grote, M.N., Venti, J., Borgerhoff-Mulder, M. 2012: Cultural macroevolution on neighbor graphs : vertical and horizontal transmission among Western North American Indian societies. *Human Nature*, 23, 283-305.

Tuffley, C. and Steel, M. 1997: Links Between Maximum Likelihood and Maximum Parsimony under a Simple Model of Site Substitution, *Bulletin of Mathematical Biology,* 59, 581-607.

Vaan de, M. 2008: *Etymological Dictionary of Latin and the other Italic Languages*. Leiden/Boston: Brill Publishers.

Velasco, J.D. 2008: Species Concepts Should Not Conflict With Evolutionary History, But Often Do, *Stud. Hist. Phil. Biol. & Biomed. Sci.* 39, 407-414.

Velasco, J.D. 2012: The Future of Systematics: Tree Thinking without the Tree, *Philosophy of Science* 79(5), 624-636.

Warnow, T., Evans, S., Ringe, D., and Nakhleh, L. 2006: A stochastic model of language evolution that incorporates homoplasy and borrowing. In P. Forster and C. Renfrew (eds.), *Phylogenetic Methods and the Pre-history of Languages*. Cambridge: MacDonald Institute for Archaeological Research, 75-90.

Weiss, M. 2014: The Comparative Method. In Bowern, C. and Evans, B. (eds.), 2014: *The Routledge Handbook of Historical Linguistics*. Abingdon: Routledge, 127-145.

Wiley, E. 1975: Karl Popper, Systematics, and Classification. *Systematic Zoology*,24, 233- 242.

1. \* Special thanks to Elliott Sober, the editor, and the anonymous reviewers for helpful comments and suggestions.

**Address for Correspondence:** Department of Philosophy, University of Wisconsin–Madison, 5185 Helen C. White Hall, 600 North Park Street, Madison, WI 53706, USA [↑](#footnote-ref-2)
2. For instance, Mace and Holden (2005) write: ‘Anthropologists are fond of pointing out the complexity of cultural systems, and either using it as an excuse to not ask precise questions, or to question the validity of the assumptions of the models being used’ (p. 120). In addition, Gray et al. (2007) report that ‘[a]t a recent symposium on phylogenetic methods in archaeology, one long-term phyloskeptic barked, ‘This is not history! This is history put into nested boxes!’ ’ (p. 366). [↑](#footnote-ref-3)
3. See Baum and Smith (2012) for an overview of cladistic parsimony and other phylogenetic methods. [↑](#footnote-ref-4)
4. For a survey of some of these heuristic methods, see Baum and Smith (2012, pp. 187-192). [↑](#footnote-ref-5)
5. It should be noted that the phylogenetic project of charting the history of language change is distinct from the interdisciplinary project of determining how the human capacity to use language evolved and developed (e.g. Hauser et al. 2002). [↑](#footnote-ref-6)
6. It is a notational convention among linguists to use an asterisk to mark a reconstructed proto-term. [↑](#footnote-ref-7)
7. See Greenhill and Gray (2005, pp. 39-42) for further discussion on coding character traits. For a comprehensive database of cognate sets for the Indo-European language family, see the “Indo-European Lexical Cognacy Database” online at http://ielex.mpi.nl/ [↑](#footnote-ref-8)
8. However, one might choose to use non-dichotomous traits, and code, for example, the character state for the meaning *hand* by giving a 1 to both German and English, a 2 to French, Spanish, and Italian, and a 3 to Russian (Ringe et al. 2002, pp. 71-2). [↑](#footnote-ref-9)
9. See Nichols and Warnow (2008) for a more detailed survey of a number of recent phylogenetic studies of language. [↑](#footnote-ref-10)
10. Consistency index is one measure of the amount of homoplasy on a tree, and in a sense describes how tree-like a given tree is. It is defined as the minimum number of changes in a data set divided by the number of changes required by the tree. Perfect fit between a data set and a tree is achieved when CI=1. [↑](#footnote-ref-11)
11. See, for instance, the introduction in Durie and Ross (1996), Rankin (2003) in Joseph and Janda (2003), and Weiss (2014) in Bowern and Evans (2014) for more detailed overviews of the comparative method [↑](#footnote-ref-12)
12. In linguistics, a ‘morpheme’ is commonly defined as the ‘the smallest unit of meaning’ (Greenberg 1977, p. 51). A word can be or contain just one morpheme, but many words contain more than one. For example, the word ‘help’ contains one morpheme. However, ‘helplessness’ contains three morphemes: ‘help’ the root noun; ‘-less’, a suffix which makes a noun an adjective and denotes an absence; and ‘-ness’, another suffix which turns an adjective into a noun and denotes a state or quality of being. [↑](#footnote-ref-13)
13. For instance, Atkinson and Gray (2005, p. 513) claim that the two are ‘demonstrably analogous.’ See also Mace and Holden (2005, p. 167) for a table of analogies, and Bowern and Evans (2014, pp. 3-8) and the literature cited therein for a more recent overview. [↑](#footnote-ref-14)
14. According to Campbell (1998, pp. 115-122), rules of thumb used to infer proto-sounds include ‘economy’, where a single change is preferable to multiple changes. See Lass (1990, p. 148) who appeals to parsimony when accounting for certain grammatical peculiarities of Southern Hiberno-English, and also Lass (1997, ch. 3-4) for further appeals to parsimony and other cladistic concepts. [↑](#footnote-ref-15)
15. Another motivation for embracing linguistic phylogenetics is that it might offer a solution to the problem of language individuation. Given that the common criterion of ‘mutual intelligibility’ should preserve transitivity but does not, owing to the existence of ‘dialect chains’ (Dixon 1997, p. 8), linguists might consider instead adopting a phylogenetic concept of language, similar to the ‘Phylogenetic Species Concept’ certain theorists (e.g. Donoghue 1985, Mishler and Brandon 1987; Velasco 2008) propose as a way to deal with problems that afflict the popular ‘Biological Species Concept’, which individuates species according to inter-fertility (Mayr 1942). [↑](#footnote-ref-16)
16. Here and throughout, I will refer to language as a ‘cultural artifact’. Some might find this usage problematic, but by such usage I do *not* mean to say that the development of the human capacity for language cannot be given a biological explanation (see fn. 4). [↑](#footnote-ref-17)
17. See Boyd et al. (1997) who helpfully distinguish four different models of culture, determined by levels of coherence and stability. [↑](#footnote-ref-18)
18. See ch. 2 ‘Lexical Borrowing: Concepts and Issues’ in Haspelmath and Tadmor (2009) for a general discussion (pp. 35-54). [↑](#footnote-ref-19)
19. This follows the practice of Swadesh (1952) which introduces a 200-word list. This list is shortened to 100 in Swadesh (1955). [↑](#footnote-ref-20)
20. Indeed, it is not even clear that we can draw general conclusions about patterns of transfer in the domain of *material* culture, based solely on the negative result of Temkin and Eldredge (2007) canvassed above. A study of the evolution of Turkmen textiles by Collard and Tehrani (2005), which compares a hypothesis of vertical transmission to one of horizontal transmission, suggests a branching pattern rather than a pattern of horizontal diffusion as that which best accounts for the data (p. 125). [↑](#footnote-ref-21)
21. Towner et al. (2012) attempt to determine how tree-like elements of culture are by embedding cultural traits in neighbor graphs in order to produce ‘networks’ instead of trees via the NeighborNet algorithm, thereby allowing representation of both vertical and horizontal transmission. Different models—some of which allow no horizontal transmission—are then evaluated for their predictive accuracy using model selection criteria such as the Akaike Information Criterion and the Bayesian Information Criterion. See also Gray, Bryant, and Greenhill (2010) for a similar attempt using NeighborNet to address the debate of horizontal diffusion locally rather than globally. The findings of Gray, Bryant, and Greenhill support a tree-like model for languages in some parts of the world but not for others. [↑](#footnote-ref-22)
22. See various entries in the edited volume of Aikhenvald and Dixon (2001) for worries of this sort among linguists. Instead of a tree-like model, some contributors propose that models based on clouds, cyclones, or thickets may be more appropriate representations (pp. 4-5). [↑](#footnote-ref-23)
23. See Velasco (2012) who offers a similar defense of phylogenetic trees in biology. See also Doolittle and Bapteste (2007) who argue for a ‘pattern pluralism’, whereby different representations will be appropriate and true for different taxa at different levels of description. [↑](#footnote-ref-24)
24. See, for instance, Aunger's (2001) edited volume *Darwinizing Culture: The Status of Memetics As a Science.* [↑](#footnote-ref-25)
25. Similar to a ‘morpheme’ (see fn. 11) a ‘phoneme’ is the smallest unit of sound in a language that creates a difference in meaning. For instance, in English, the [p] as in ‘pat’ and the [b] in ‘bat’ are distinct phonemes. This is not true in every language though, e.g. Korean. [↑](#footnote-ref-26)
26. However, see Mace and Holden (2005, p. 167) in which this assumption is acknowledged. In the cited table showing the parallels between biology and culture, ‘cultural traditions, memes, ideas, artifacts, words, grammar and syntax’ are treated as discrete units. [↑](#footnote-ref-27)
27. See also Sterelny (2006) who argues, in response to the intense focus on memes that there are other kinds of mechanisms of inheritance available. Both niche construction acting at the level of groups (e.g. fire-making, pot-making, or any technological advance) and vertical transfer of information—though not replication—from parent to offspring are viable mechanisms that would lead to evolution. See also Heinrich et al. (2008), for dissatisfaction with the undue focus on meme-like inheritance in discussions of cultural evolution. [↑](#footnote-ref-28)
28. Darwin himself was cautious about issues of inheritance, did not know about Mendel’s work on genes, and furthermore, held ideas about the nature of inheritance in his (1868) that turned out to be at odds with Mendel’s particulate theory of inheritance. [↑](#footnote-ref-29)
29. The principle according to which only neutral and deleterious traits are evidence of common ancestry, Sober calls ‘Darwin’s Principle’ (2009, p. 10051). While there are exceptions to the principle—for example, when an adaptive trait is correlated with a neutral trait—legitimate applications of Darwin’s Principle can be given a straightforward probabilistic justification, given that it is often the case that Pr(two species share a non-adaptive trait | common ancestry) > Pr(two species share a non-adaptive trait | separate ancestry). [↑](#footnote-ref-30)
30. See Sober (1988, ch. 4; 2015, ch. 3) for a critical survey of some other attempts to justify the use of cladistic parsimony, some of which include methodological necessity (Hennig 1966), the a priori assumption that nature is simple (Camin and Sokal 1965), dubious appeals to Popperian falsificationism (Eldredge and Cracraft 1980), and avoiding ad hoc hypotheses (Farris 1983). While Sober is critical of Farris (1983), it is not clear that his critique applies to the ‘Inference to the Best Explanation’ justification that I sketch in section 7, although it should be noted that Sober is no friend of Inference to the Best Explanation (Roche and Sober 2013). [↑](#footnote-ref-31)
31. In addition, since the Tuffley-and-Steel model allows each site to evolve according to its own rules, and thus potentially contains a vast number of adjustable parameters, it scores extremely poorly according to model-selection criteria such as the Akaike Information Criterion (AIC), which imposes a penalty for complex models (Akaike 1973). See Holder, Lewis, and Swofford (2010) for a discussion. See Forster and Sober (1994) for an accessible overview of AIC and its epistemological significance. [↑](#footnote-ref-32)
32. Warnow et al. (2006) uses a variant of the Tuffley-and-Steel model, which therefore assumes that characters evolve independently. The assumption of character independence might not prove problematic provided one chooses characters known to evolve by drift. But even seemingly neutral grammatical or syntactic features can be subject to prestige bias or other social forces. For instance, Pawley and Syder (1983) argue that a number of grammatical and syntactic features that have arisen in modern English can be explained in terms of natural selection, e.g. the proliferation of the passive voice given the rise of science and the need for portraying impersonal objectivity. [↑](#footnote-ref-33)
33. See Kluge and Grant (2006, pp. 280-1) for a critique of Farris (1983). In addition, see Farris (2008, pp. 4-6) for a response in which it is argued that Kluge and Grant (2006) aren’t saying anything substantively different from Farris (1983). In the main, their disagreement turns on the way in which character states are conceptualized. [↑](#footnote-ref-34)
34. The ‘P’ in EPP is for ‘Popper’, as other probabilistic measures of explanatory have been put forth, as I will discuss below. [↑](#footnote-ref-35)
35. As a justification for this principle, Nolan (1997) and Baker (2003) draw inspiration from its apparent deployment in the β-decay/ neutrino controversy in particle physics in the early 20th century. [↑](#footnote-ref-36)
36. What sort of conclusion can be justified on the basis of IBE and its relationship to the most popular approach to theory confirmation, Bayesianism, is a complicated matter. See Cabrera (2015) for a discussion. [↑](#footnote-ref-37)
37. Despite not explicitly defending IBE, Quine and Ullian (1970) are an important, early source for understanding theory-evaluation by reference to a plurality of virtues. Indeed, so is Kuhn (1977), a fact which is often neglected in discussions of IBE. [↑](#footnote-ref-38)
38. Brössel (2013, p. 47) [↑](#footnote-ref-39)
39. While one might consider this a problem for IBE, it’s not clear whether having a formal, probabilistic measure of explanatory power is desirable or even possible. See Glymour (2015) for a critique of this increasingly popular research program. [↑](#footnote-ref-40)
40. As is well-known, Felsenstein (1978a) first highlighted the problem of ‘long-branch attraction’, which refers to circumstances in which evolutionary change is accelerated, and as a result two lineages will be mistakenly determined to be more closely related than they in fact are by maximum parsimony. Thus, in some circumstances parsimony will be inconsistent, i.e. the method will not converge with probability one on the correct tree as more and more data is gathered. Felsenstein (1978a) claims the advantage of maximum likelihood is its avoiding this problem (p. 408). However, this claim was shown to be mistaken, as even maximum likelihood can be inconsistent, if, for instance, the underlying model is false (Farris 1983, 1999). What’s more, assuming the model of Tuffley and Steel (1997), maximum likelihood can be inconsistent even if the model is true (p. 597). Given that maximum likelihood affords no advantage over parsimony vis-à-vis the problem of statistical inconsistency, in recent years, discussion concerning the contest between likelihood and parsimony has ceased to be centered on consistency (Goloboff 2004, p. 93). [↑](#footnote-ref-41)
41. As Sober (2015, p. 209) suggests the assumptions might vary from problem to problem, and so talk of ‘the’ assumptions is misleading. [↑](#footnote-ref-42)